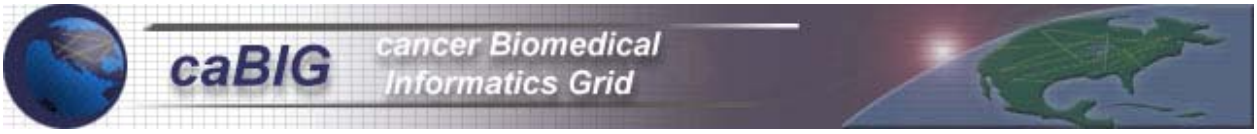

Integrative Cancer Research Special Interest Group Teleconference

Genome Annotation (formerly Gene Annotation) SIG Meeting Minutes

Date, Time & Location:	June 3, 2004 2:00 – 3:00 EDT
Attendees:	Cathy Wu – Georgetown Rakesh Nagarajan – Wash U Craig Street – Penn Harold Riethman – Wistar Yubarsud Narasimhan – Center for Cancer Research Lincoln Stein – Cold Spring Harbor Terry Disz - U of C Ross Overbeek – U of C Veronika Vonstein – U of C Yajun Yi - Vanderbilt Claire Zhu – BAH Juli Klemm - BAH
Introduction:	**Roll-call, open meeting, review meeting goals** - Review of last meeting - Review mission statement - Review Developer/Adopter activities - Identify and define future activities/research areas
Review Discussions:	**Review discussion of last meeting** - Computational genomics SIG has been combined with Gene Annotation SIG. - The group discussed the definition of gene annotation, and agreed that it should encompass structural and functional information, as well as any types of descriptive information associated with genes and genomes such as SNP, haplotype, etc. It was suggested at this meeting that the combined SIG adopts the name “Genome Annotation SIG”. - The group discussed instability and ambiguity in gene identifiers and how to deal with such problems. **Review of Mission Statement** - Harold Riethman (Wistar) brought up that this group should keep abreast of large annotation projects like ENCODE, a trans-NIH project to improve the annotation of the human genome. o In general, the group agreed that we should be keep up to date with key annotation efforts. One recommendation is to put together a compendium of these efforts and to report on them with some frequency. - Juli pointed out that although the effort of the current year is focused on existing tools, caBIG is a multi-year project and the mission statement should capture long-term goals. - There was a question about status of caBIG APIs. These are still being defined by the architecture group; caCORE can be considered a starting point. It was also brought up that training of developers on caBIG APIs as they become available is a very important issue. Coordination with the training working group will be

	necessary.
	
	Review of Developer/Adopter activities
	
	- Juli has been contacting centers individually and discussed resources and timelines with each centers as part of the matchmaking process. There are verbal understandings of the projects to be undertaken and developer/adopter pairings for these. This process has been focused on funded efforts/interests, but unfunded Adopter roles are highly encouraged.
	- U of Iowa Holen will adapt TrAPSS to the caBIG architecture; The Institute for Cancer Prevention will adopt this tool
	- Georgetown will make PIR data available to caBIG; Penn will be the adopter for this project
	- Wash U will likely adapt Function Express to the caBIG architecture; an adopter needs to be identified. Rakesh added that a separate database has been created for caBIG usage, and will be ready for distribution in a week or so. It will be best to do a demo to the group prior to distribution.
	- The Center for Cancer Research will adapt GOMiner to the caBIG architecture; Wistar is the potential adopter for this tool.
	- The Burnham Institute will adapt Cancer Molecular Pages to the caBIG architecture; The Institute for Cancer Prevention will adopt this tool
	- Cold Spring Harbor will make GKB, HapMap, and PromoterDB data available to caBIG; Wistar and Sloan will be adopters for these projects.
	- University of Chicago (Argonne) will adapt SEED to the caBIG architecture; an adopter needs to be identified. The Argonne group gave an overview of SEED:
	  o Designed for comparing "pathways" across genomes
	  o Has some automatic annotation function, but the main feature is to support manual curation of genes across a spectrum of genomes.
	  o Contains 200-300 genomes – mostly prokaryotes and some eukaryotes
	  o Support gene annotation with protein-based evidences as well as literature.
	  o Pseudogenes and frameshifts are not explicitly handled, though comparisons across multiple genomes can help identify these issues.
	  o There may be a compliment between Georgetown's UniProt effort and SEED -- SEED may be appropriate to support PIR's protein family-based approach to annotation..

Future Activities	**Current projects**
	
	- Presentation/demo on tools by Developers on the next few meetings
	  o ~ 20 min presentation/demo (PowerPoint, live demo) on each tool, followed by ~ 10 min discussions.
	  o For the next meeting, Lincoln Stein will give a demo of GMOD tools, Terry will do a demo of SEED. Juli will follow up on format and providing resources.
	- In the future, this meeting will be used to update on ongoing projects, and to

resolve issues and problems.

Future research areas

- Identify mechanism for keeping updated with current projects
- Should caBIG utilize LSID? Lincoln Stein gave an introduction on LSID:
  - Stands for Life Sciences Identifier, a type of URN.
  - 2-3 years into development by I3C, a consortium of Biotech and Pharmas.
  - Is a global naming system for biological objects ranging from patient samples to genes.
  - Naming is a big problem when it comes database integration. LSID ensures unique names for any biological objects.
  - Takes the form: urn:lsid:AuthorityID:NamespaceID:ObjectID:RevisionID
  - Namespace owned by each organization. No need for registration.
  - CSH has used LSID extensively for the haplotype mapping project to keep track of SNPs and individual genotyping results. One drawback of the LSID is that the names become very long – not "friendly on the eye."
  - The are a few reference implementations of resolution services available for LSIDs – both IBM and the Broad Institute sponsor such services.
  - For caBIG, the general decision will be between a URL or a URN-based approach to identifiers.

Other Items Discussed

- Time for future meetings will be changed from 2:00 PM to 3:00 PM ET to accommodate more participants.

Action Items:

Name Responsible	Action Item	Date Due	Notes
Juli Klemm	Distribute meeting minutes	6/7/04	
Terry Disz, Lincoln Stein	Present SEED and GMOD at next month's SIG meeting	7/1/04	
Juli Klemm	Follow up with Lincoln and Terry on presentations/demo	6/21/04	
Juli Klemm	Follow up with Yajun Yi	6/11/04	Regarding minimizing duplication of efforts within the SIG
All Participants	Genome Annotation Compendium	TBD	